J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 5257--5275

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

# Spatio-temporal Evolution and Influence Mechanism of Rural Tourism Flows Based on Network Data Mining

Tingting Jin[1,✉]
[1] Boda College of Jilin Normal University, Siping, Jilin, 136000, China

ABSTRACT
Rural tourism, as an important part of the tourism service industry, the study of the spatio-temporal evolution and influence mechanism of rural tourism flow has also become a hot topic at present. This paper takes Jiangsu Province of China as the research area, proposes the heat measurement and identification method of rural tourism based on network data, constructs a heat measurement model, takes standard deviation ellipse analysis, average nearest neighbor index method, kernel density analysis as the core method of spatial analysis, and proposes the hotspot identification method on the demand of spatial relevance analysis, so as to provide the method and means for the analysis of the spatio-temporal evolution of the rural tourism flow. In the analysis of the influence mechanism of rural tourism flow, the QAP model is used as a research tool to explore the influencing factors of rural tourism flow.The value of rural tourism hotness was low during 2014-2017, and it has rapidly increased and maintained a high growth trend since 2018. The Gini coefficient of rural tourism hotness increased from 0.51 in 2014 to 0.72 in 2018, and then fell back to 0.65 in 2023, and the degree of spatial difference of rural tourism hotness showed a weakening trend, and the hotspot areas of rural tourism were increasing. The structure of tourism flow is affected by a variety of factors such as spatial proximity, tourism income, and the impacts produced by the factors change somewhat in different time periods.

*Keywords:* heat measurement model, hotspot identification, spatial analysis, QAP model, rural tourism flow, spatio-temporal evolution

## 1. Introduction

Rural tourism is a new type of tourism form with the purpose of tourism and vacation, with the village wilderness as the space, with the humanities without interference, ecology without destruction, with the characteristics of the tour of residence and the village wilderness act of the village wilderness

✉Corresponding author.
*E-mail address:* 13634464978@139.com (T. Jin).
Received 10 January 2024; Revised 15 March 2024; Accepted 25 December 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-296

tourism form, which is a new type of tourism form [1]. At present, urban residents work under greater pressure, in order to ease the daily pressure, people are more eager to return to the rural life. Rural tourism is to seize people's psychology, the “three rural” resources as a tourism carrier, rural culture, natural landscape, rural scenery, agricultural production and agricultural products as the main attraction, to provide travelers with a natural, leisure and healthy experience of rural life tourism [2-4]. This provides a new growth point for the tourism economy and an effective carrier to reasonably balance the rural industrial restructuring, farmers' well-being and urban residents' multiple needs for leisure tourism [5-6]. It plays a positive role in effectively solving the “three rural issues”, integrating urban and rural development and promoting the transformation and upgrading of tourism [7-8].

In the new era of China's comprehensive promotion of rural revitalization, rural tourism is facing new development environment and opportunities, but also new problems and challenges. The expansion of the scale of tourists and the extensive use of Internet information technology promote changes in the form of tourism consumption, tourism organization, and characteristics of tourism demand, which together drive the change of tourism industry [9-11]. In order to adapt to the new changes occurring in the tourism market and meet the quality, personalization and diversification needs of tourists, tourism destinations have actively transformed the development mode, improved the quality of services, and realized the transformation of the tourism development mode from rough and inefficient to high-quality development [12-15]. In addition, with the expansion of tourism scale, the transformation of tourism organization mode and the improvement of transportation conditions, the tourism industry organization network will continue to expand, and more tourist destinations, sources and transit places will be connected, making it easier to form a polycentric tourism flow spatial network structure [16-19]. The study of rural tourism flow helps to identify the behavioral characteristics of rural tourists, recognize the spatial and temporal flow patterns of rural tourism flow, identify the hot spots of rural tourism flow, reveal the influence mechanism of rural tourism flow, and is of great significance to the regional cooperation and coordinated development of rural tourism, spatial optimization and traffic diversion, and planning, construction, and marketing, etc. [20-23].

This paper focuses on the spatio-temporal evolution and influence mechanism of rural tourism flow, in order to realize the deepening of rural tourism flow research content and perspective innovation at the theoretical level. In the study of spatio-temporal evolution of rural tourism flow, the attraction heat measurement model is proposed, based on the tourism website data, the heat of rural tourism attractions is divided into evaluation heat and attention heat, and the heat of rural tourism attractions is reflected through the attention of potential tourists. Standard deviation ellipse analysis, average nearest neighbor index analysis and kernel density analysis are selected as the main methods for rural tourism spatial evolution analysis. The standard deviation ellipse analysis method reflects the contour and dominant direction of geographic elements in spatial distribution as a whole, and the average closest neighbor index analysis method and nuclear density analysis method are used to calculate the closest neighbor index of the region and the density of spatial elements in the surrounding neighborhood, respectively. In the analysis of the influence mechanism of rural tourism flow, QAP model was introduced to establish QAP weighted network regression equation and calculate QAP correlation coefficients with matrix network as the explanatory and interpreted variables. Combined with the network data mining, the spatial and temporal evolution of rural tourism flows and the in-depth analysis of the influence mechanism are carried out with Jiangsu Province, China, as the study area.

## 2. Data sources and processing

### *2.1. Selection of the study area*

This paper takes Jiangsu Province in China as the study area, and the main research units involved include rural tourism scenic spots, provincial areas and county areas.

Customarily, Jiangsu Province is geospatially divided into three parts: southern Jiangsu, central Jiangsu, and northern Jiangsu: southern Jiangsu includes the boundaries of the five prefecture-level cities of Nanjing, Zhenjiang, Changzhou, Wuxi, and Suzhou, which are south of the Yangtze River; central Jiangsu usually refers to the boundaries of the three prefecture-level cities of Yangzhou, Taizhou, and Nantong, which are on the north bank of the Yangtze River; and northern Jiangsu includes the boundaries of the five prefecture-level cities of Xuzhou, Huai'an, Yancheng, Lianyungang, and Suqian.

Since rural tourism sites are usually not located in the central urban area of each prefecture-level city, this paper combines each central urban area of a prefecture-level city with the suburban areas surrounding the central urban area into one geographic unit with the name of prefecture-level city urban area (e.g., Jingkou, Runzhou, and Dantu districts of Zhenjiang City are combined into one county-level geographic unit called Zhenjiang urban area). The relatively independent suburban districts and counties, county-level cities as independent county-level geographical units. The 98 county-level administrative divisions in Jiangsu Province are thus divided into 64 county-level units.

### *2.2. Data sources and description*

Network travelogue is one of the important ways to characterize tourism big data, including photos, videos and texts with geographic information location, etc. It has become an important data source for scholars to study tourism flow because of its features of large data volume, wide coverage, large time span, easy access, etc., which can more accurately and realistically reflect the routes of tourists' flow and spatial and temporal characteristics. The website of “Webmaster Home” (https://www.chinaz.com/) mainly ranks mainstream websites in various industries based on Alexa ranking, Baidu weight, PR value and other indicators, and has strong authority. In order to ensure the reliability and representativeness of the online travelogue, this study refers to the relevant literature and screens the travel websites according to the ranking of travel websites on the Webmaster's Home website, while taking into account the interface settings of each travel website, its main business, and the composition of the data content of the online travelogue, etc., and ultimately selects Ctrip, Ma's Nest, and Where to Go as the data collection sources for this study. Ctrip.com, MaHoneycomb.com, and GoWhere.com were finally selected as the data collection websites for this study, and the data on the origin, destination, departure time, and travel routes posted by tourists from 2012-2022 were collected from the above three specialized travel websites. Web data are characterized by complexity, commercial manipulation, and difficulty in identifying authenticity. Therefore, it is necessary to scientifically screen the original web travel data to meet the research requirements. The column of outdoor sports tourism keywords is set when crawling data, and the crawled network travelogues are presented in the form of EXCEL table, and the data that do not contain outdoor sports tourism keywords are excluded through the filtering function in the EXCEL table, and finally 184,866 data containing outdoor sports tourism keywords are obtained.

## 3. Spatial and temporal evolution of rural tourism flows based on network data mining

Rural tourism flow flow is an important characterization to measure the quality of rural tourism development and regional differences. Taking Jiangsu Province of China as the study area, this paper proposes a rural tourism heat measurement method based on network data mining to study the spatial and temporal evolution characteristics of rural tourism flows.

### *3.1. Rural Tourism Heat Measurement and Hot Spot Identification Methods*

3.1.1. Heat measurement models. Heat degree, originally refers to the degree of attention and demand for popular products. Scenic network heat refers to the network search, check-in, favorites, likes and other network data information to reflect an attraction in the degree. Scenic spot network heat refers to the network search, check-in, favorites, likes and other network data information to reflect an attraction in a period of time network evaluation and the degree of attention. This paper constructs a rural tourist attraction heat measurement model, which innovatively divides rural scenic spot heat into evaluation heat and attention heat according to the heat information presented by different behaviors of tourists in the tourism website before and after the tour. The main indicators of evaluation heat include the number of attraction travelogues and the number of attraction reviews, which reflect the attraction heat through the quantity and quality of network evaluation of the attraction by the visited tourists, and the main indicators of attention heat include the number of travelogues browsed and the number of attraction Q&A, which reflect the attraction heat through the attention of the potential tourists with the following formulas:

$$H_i = \sum_{j}^{12} (V_{ij} + A_{ij}) \tag{1}$$

Where, $H_i$ denotes the hotness of the $i$ attraction, $V_{ij}$ denotes the rating hotness of the $i$ attraction in year $j$, and $A_{ij}$ denotes the attention hotness of the $i$ attraction in year $j$:

$$V_i = w_1 N_i P_n i + w_2 N_1 C_i P_{ai} \tag{2}$$

Where, $V_i$ denotes the evaluation heat of $i$ attractions, $N_i$ denotes the number of travelogues of $i$ attractions. $P_{ni}$ denotes $i$ the quality of the attraction's travelogues noting the number of likes, the number of travelogue reviews linearly normalized to the weight coefficients in the range of [0, 1], with the specific formula $P_{ni} = 1 +$ normalized number of travelogue likes + normalized number of travelogue reviews). $C_i$ denotes $i$ the number of attraction reviews. $P_{ci}$ denotes $i$ attraction review quality (normalizing the number of review likes. The number of review comments is linearly normalized to a weighting factor in the range of [0, 1], with the same formula as the quality of travelogues). $w_1, w_2$ is the weight of the number of travelogues and the number of reviews:

$$A_i = w_3 B_i + w_4 Q_i P_{qi} \tag{3}$$

In the formula, $A_i$ denotes the attention heat of $i$ attractions, $B_i$ denotes the number of tour browsing of $i$ attractions, $Q_i$ denotes the number of Q&A of $i$ attractions, $P_{qi}$ denotes the Q&A quality of $i$ attractions (the number of Q&A browsing, the number of Q&A likes, and the number of Q&A comments are linearly normalized to the weighting coefficients in the range of [0, 1], with the same formula as that for the quality of the tour), and $w_3, w_4$ is the weighting of the number of tour browsing and the number of Q&A.

3.1.2. Methods of spatial analysis:

1) Standard deviation ellipse analysis. The standard deviation ellipse is to reflect the contour and dominant direction of geographic elements in spatial distribution from the whole, and the distribution range of standard deviation ellipse is the main range of geographic elements in spatial distribution [24]. The standard deviation ellipse mainly expresses the evolution of spatial and temporal patterns of rural tourist attractions in Jiangsu Province through the coordinates of the center of the ellipse, and the standard deviation $\sigma_x$ and $\sigma_y$ of the standard deviation ellipse's corners $\theta, x$ and $y$ axes. The center coordinates indicate the relative position of the distribution of rural tourist attractions in Jiangsu Province in two-dimensional space, the corner $\theta$ indicates the trend and direction of the spatial distribution of rural tourist attractions in Jiangsu Province, the long axis of the ellipse is the direction of the most spatial distribution of rural tourist attractions in Jiangsu Province, and the short axis of the ellipse is the direction of the least spatial distribution of rural tourist attractions in Jiangsu Province. Its formula is:

$$\tan\theta = \sum_{i=1}^{n}(x_i-\bar{x})^2 - \sum_{i=1}^{n}(y_i-\bar{y})^2$$
$$+\left[\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 - \sum_{i=1}^{n}(y_i-\bar{y})^2\right]^2 + \cdots + \frac{1}{2}(x_i-\bar{x})^2 - \sum_{i=1}^{n}(y_i-\bar{y})^2\right]^2 \tag{4}$$
$$+4\left[\sum_{i=1}^{n}(x_i-\bar{x})\sum_{i=1}^{n}(y_i-\bar{y})\right]^{\frac{1}{2}} / 2\sum_{i=1}^{n}\sum_{i=1}^{n}(x_i-\bar{x})\sum_{i=1}^{n}(y_i-\bar{y})$$

$$\sigma_x = \sqrt{\sum_{i=1}^{n}[(x_i-\bar{x})\cos\theta-(y_i-\bar{y})\sin\theta]^2/n} \tag{5}$$

$$\sigma_y = \sqrt{\sum_{i=1}^{n}[(x_i-\bar{x})\sin\theta-(y_i-\bar{y})\cos\theta]^2/n} \tag{6}$$

where $(x_i, y_i)$ denotes the geographic coordinates of rural tourist attractions in Jiangsu Province and $\theta$ denotes the azimuth of the ellipse.
2) Average closest neighbor index analysis method. The closest neighbor index refers to the ratio of the actual closest neighbor distance and the theoretical closest neighbor distance [25]. First, the Hawths Tools in ArcGIS 10.3 software is used to measure the Euclidean distance between rural tourist attractions in Jiangsu Province and their closest neighboring attractions, and then the average value, which is the actual closest distance, is calculated, and then the theoretical closest distance is calculated with the formula:

$$\overline{r_E} = \frac{1}{2\sqrt{n/A}} = \frac{1}{2\sqrt{D}} \tag{7}$$

Where, $A$ represents the area of Jiangsu Province, $n$ represents the number of rural tourist attractions in Jiangsu Province, and $D$ represents the density of rural tourist attractions. Finally, the closest neighbor index is calculated with the formula:

$$R = \frac{\overline{r_1}}{r_E} = 2\sqrt{Dr_1} \tag{8}$$

3) Kernel density analysis method. Kernel density refers to the density of spatial elements in the surrounding neighborhood, reflecting the distribution characteristics of geographic phenomena [26]. Taking the position of a specific element point as the center, the attributes of the element point are distributed within a specified threshold, the density of the center position is maximum and decays with distance, and the density is zero at the limit distance. The formula is:

$$f_n(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x-x_i}{h}\right) \tag{9}$$

where $f(x)$ is the kernel density estimate, known as the kernel function, $h$ is the bandwidth, whose value is greater than zero, and $x-x_i$ is the distance from valuation point $x$ to data point $x_i$.

3.1.3. Hot spot identification methods:
1) Calculation of hotspot analysis. Hot spot analysis is one of the spatial correlation analysis, this analysis method is mainly on the need to analyze the elements of the dataset to calculate Getis-dGi statistics (known as $G_i^*$, return $G_i^*$ and statistics $Z$ scores, can also be called statistical weighted significance of the hot spot and cold spot analysis, and its calculation formula is shown in (10) [27]:

$$G_i^{å} = \frac{\sum_{n=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j} - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}} \tag{10}$$

Eq. (10) where $G_i^*$ is the statistical $Z$ score, $\bar{X}$ is the mean value of the attributes of the elements in the dataset, S is the standard deviation of each element, $x_j$ represents the attribute values of the $j$ elements, usually indicating attributes such as distance, area, score, etc., and $w_{i,j}$ represents the relative weights between the $i$ and $j$ elements, and $n$ is the total number of elements involved in the analysis. Where $\bar{X}$ and S are calculated as in equations (11) and (12):

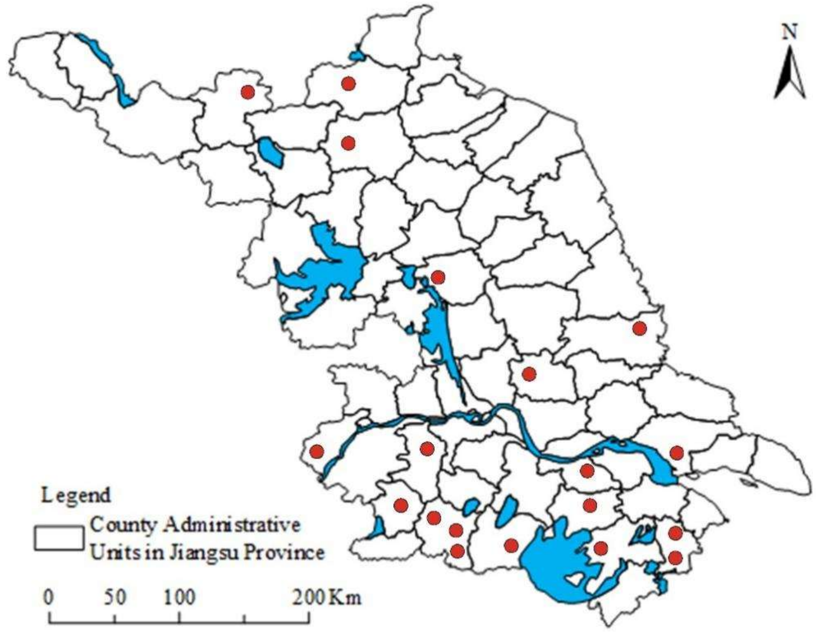
$$\bar{X} = \frac{\sum_{j-1}^{n} x_j}{n} \tag{11}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \tag{12}$$

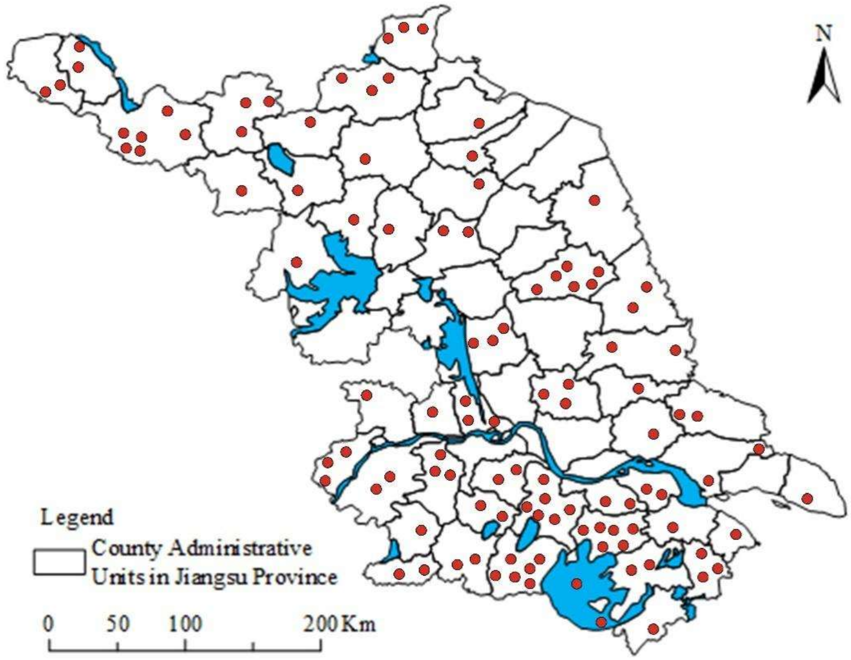
The hotspot analysis tool is to analyze the degree of agglomeration and confidence value of each element in the data set on the basis of weighting the weights of different elements, the difference between hotspot analysis and high/low cluster analysis is that the hotspot analysis does not necessarily have statistical significance, but only indicates that the high value (hotspot) is more attractive when compared with other elements in the data set, and when hotspot analysis occurs locally it has a statistically significant clustering significance.

### *3.2. Spatial distribution characteristics of rural tourist attractions*

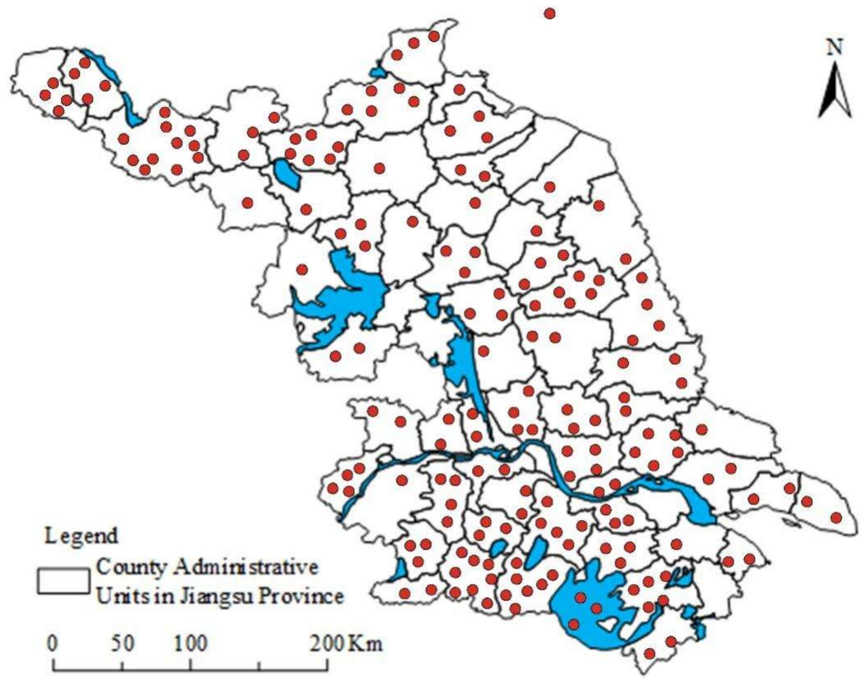
1) The number of scenic spots increases year by year, but the spatial difference is significant. The spatial distribution pattern of star-rated rural tourism areas in Jiangsu Province from 2014 to 2023 is shown in Figure 1. In 2014, there were only 19 star-rated rural tourism scenic spots in Jiangsu Province, and in 2023, the number increased to 289, an increase of nearly 15 times in ten years. The distribution pattern of rural tourism attractions from scattered scattered development to cover half of the provincial scope, the spatial and temporal distribution pattern shows year by year expansion trend, the degree of regional differentiation intensified. The overall performance of two characteristics, first, the number of rural tourist attractions increased significantly, and second, the spatial distribution of rural tourist attractions in the north and south is significantly different.


N
Legend
County Administrative Units in Jiangsu Province
0 50 100 200 Km

(a)2014


N
Legend
County Administrative Units in Jiangsu Province
0 50 100 200 Km

(b)2018




(c)2023

**Fig. 1.** The spatial distribution pattern of rural tourism areas in Jiangsu Province

Changes in the number of sub-regional rural tourism areas in Jiangsu Province in 2014, 2018 and 2023 are shown in Table 1.The number of star-rated rural tourism scenic spots in Jiangsu Province in 2023 has increased by 272 compared with that in 2014, which has covered more than 90% of the county-level administrative areas in the province. The spatial distribution of rural tourist attractions varies significantly in each period.In 2014, there were 24 star-rated rural tourist attractions, of which 15 were distributed in the southern part of Jiangsu Province, and there were only 9 rural tourist attractions distributed in the northern and central parts of Jiangsu Province.The percentage of the number of rural tourist attractions in the central and northern parts of Jiangsu Province increased in 2023, increasing to 47.26%.In 2023, with 27.16% of the land area, the southern part of Jiangsu Province clusters and distributes 55.74% of the province's rural tourist attractions. This indicates that, as an economic phenomenon, the distribution of rural tourist attractions is positively correlated with the level of economic development, i.e., the higher the level of economic development, the more potential tourists, the more market-driven, and the greater the number of rural tourist attractions.

**Table 1.** Changes in the number of rural tourist areas

| Year | Sunan area | | Suzhong area | | Subei area | |
|---|---|---|---|---|---|---|
| | Number of scenic spots | Proportion(%) | Number of scenic spots | Proportion(%) | Number of scenic spots | Proportion(%) |
| 2014 | 15 | 62.50% | 5 | 20.83% | 4 | 16.67% |
| 2018 | 78 | 50.32% | 32 | 20.65% | 45 | 29.03% |
| 2023 | 165 | 55.74% | 51 | 17.23% | 80 | 27.03% |

2) Scenic area coverage expansion, spatial distribution of the center of gravity to the north west migration. Using ArcGIS10.2 software to generate the standard deviation ellipse of the spatial

distribution of rural tourism scenic spots, drawing the center of gravity moving trajectory, specifically as shown in Figure 2.

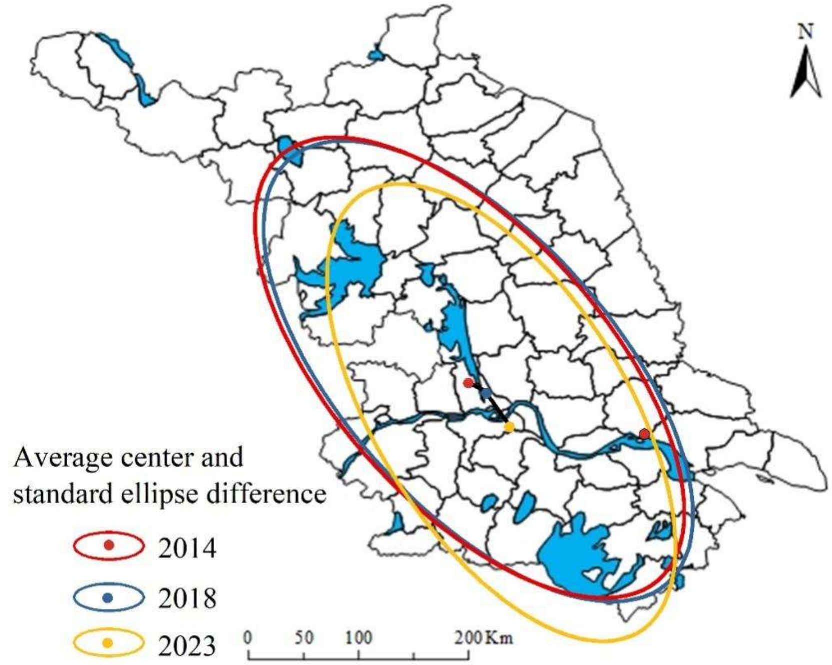



**Fig. 2.** Spatial distribution center of gravity

Meanwhile, the coordinates of the center of the ellipse, the distance moved by the center of gravity, and the parameters of the ellipse area and angle of rotation are calculated for the three time sections in 2014, 2018, and 2023, respectively, as shown in Table 2.

**Table 2.** Mass change and standard deviation elliptic parameter

| - | Gravity parameter | 2014 | 2018 | 2023 |
|---|---|---|---|---|
| Mass change | Central coordinates | 119.760°E,32.192°N | 119.521°E,32.476°N | 119.512°E,32.485°N |
| | Moving direction | - | Northwest | North by west |
| | Moving distance | - | 39.66km | 12.25km |
| Standard deviation elliptic parameter | Elliptic area | 5.141 | 5.611 | 5.749 |
| | Angle of Angle $\theta$ | 146.052 | 138.04 | 139.304 |
| | Along the x axis standard deviation (km). | 0.901 | 0.933 | 0.914 |
| | Along the Y-axis standard deviation (km). | 1.793 | 1.927 | 1.983 |
| | flattening | 0.494 | 0.518 | 0.528 |

In terms of the center of gravity of the spatial distribution of rural tourist attractions, the center of gravity of the spatial distribution of rural tourist attractions varied between 119.760°E~119.512°E and 32.192°N~32.485°N during the study period, with the center of gravity located in the city of Yangzhong in 2014, and shifted northward to the downtown area of Yangzhou in 2018. The center of gravity moves in the north-west direction, with a distance of nearly 50 km during the ten-year period, and the distance shifted in the north-south direction is larger than that in the east-west direction.The standard deviation ellipse area of the three time sections shows a yearly growth trend, which indicates that the coverage of the star-rated rural tourist attractions is expanding year by year. Overall, the cluster area of rural tourist attractions moves gradually from the southern region of Suzhou to the northern region of Suzhou, with a larger change during 2014-2018 and a smaller change during 2018-2023. The spatial distribution characteristics of rural tourist attractions in Jiangsu Province generally show a southeast-northwest pattern in terms of the range of changes in Corner $\theta$. During the period of 2014-2023, the oblateness of the standard deviation ellipse gradually increases, indicating that the directional trend of spatial agglomeration of rural tourist attractions is becoming more and more obvious, and that the spatial distribution pattern of southeast-northwest will not change much in the coming period of time.

### *3.3. Characteristics of the spatial and temporal evolution of rural tourism enthusiasm*

3.3.1. Temporal evolution of rural tourism heat. The annual change of rural tourism heat value is shown in Figure 3. During the study period, the value of rural tourism in Jiangsu Province showed an "S" type trajectory of change, and the trend of rural tourism heat increasing year by year is very obvious.During 2014-2017, the development of rural tourism was slower, and the value of rural tourism heat was lower, and the growth rate was slow, although it had been increasing. The value of rural tourism heat increased rapidly in 2018, and has been maintaining a high growth trend during 2018-2020; it began to slow down in 2016, and the heat of rural tourism in Jiangsu Province entered a stable growth stage. It is not difficult to find that the trajectory of rural tourism heat is basically consistent with the stage of rural tourism development.

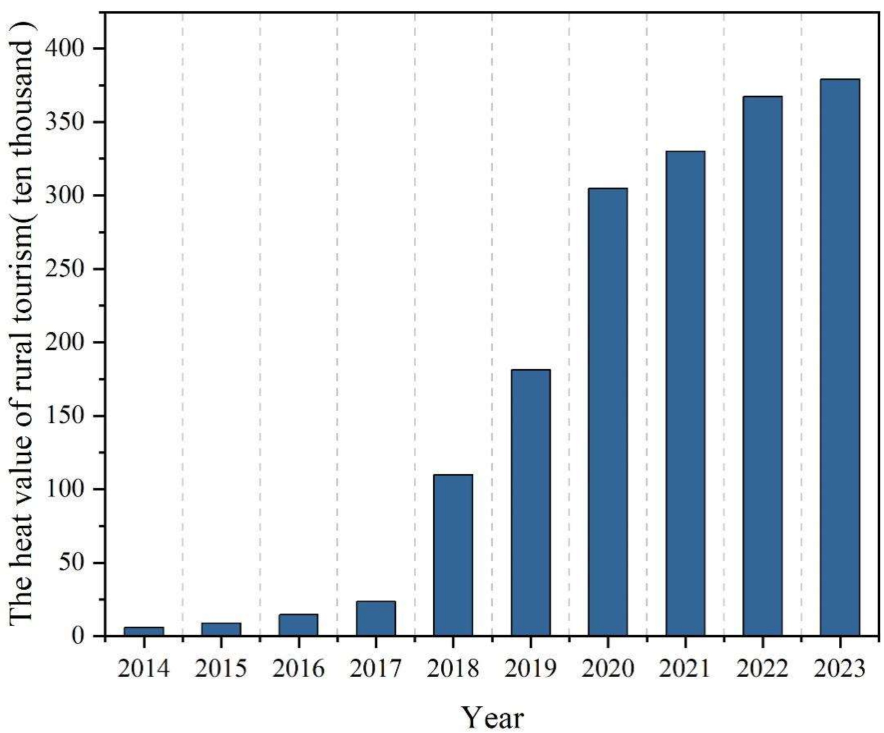



**Fig. 3.** Annual changes in the heat value of rural tourism

3.3.2. Spatial heterogeneity of rural tourism enthusiasm：

1) Spatial heterogeneity of rural tourism hotness. This paper expresses the spatial heterogeneity characteristics of rural tourism heat by calculating the Gini coefficient of rural tourism heat value, and the results are shown in Figure 4. Overall, from 2014 to 2018, the Gini coefficient of rural tourism heat value is above 0.5, which indicates that the regional distribution of rural tourism heat in Jiangsu

Province is not balanced, and the geographic clustering phenomenon of rural tourism flow is significant. Looking at the overall trend, the Gini coefficient of rural tourism heat increases from 0.51 in 2014 to 0.72 in 2018, and then falls back to 0.65 in 2023, showing an “inverted U”-type evolutionary trajectory, i.e., the degree of agglomeration first grows rapidly, and then slows down, and the Gini coefficient appears to be at its maximum, with the emergence of the inflection point of the Gini coefficient. With the appearance of the inflection point of Gini coefficient, the degree of agglomeration began to decline, that is, the spatial differences in the degree of rural tourism heat appeared a weakening trend.

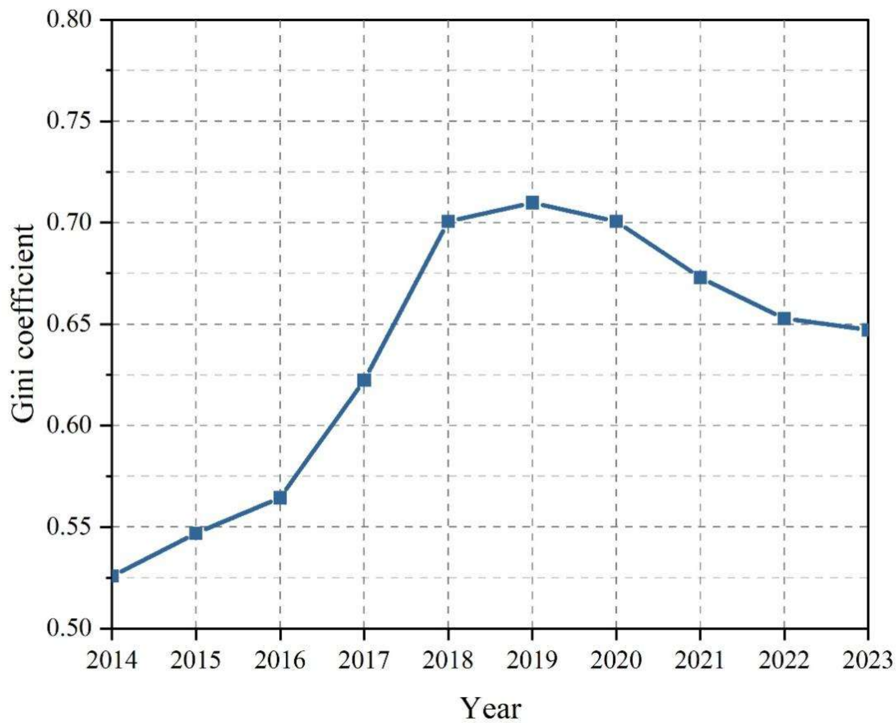



**Fig. 4.** The Gini coefficient of rural tourism heat value

2) Patterns and evolution of hot and cold rural tourism flows. Taking the county as the evaluation unit, this paper uses the hotspot analysis method to accurately identify the high-value agglomeration areas and low-value agglomeration areas of rural tourism flow, i.e., "hot spot areas" and "cold spot areas". Firstly, the Getis-Ord Gi* values of each county in 2014, 2018 and 2023 were calculated, and then the Gi* values of each year were divided into four categories from high to low by using Jenks's "natural discontinuity classification method", and the distribution map of "hot spots" and "cold spots" of rural tourism flow at the county level in Jiangsu Province was generated, as shown in Figure 5.

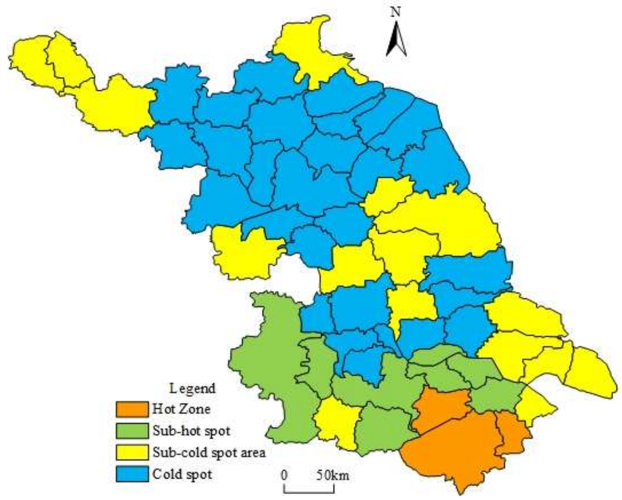
(1) Evolution of hot and cold patterns

In terms of quantitative changes, with the arrival of the rural tourism boom, the number of rural tourism hotspot areas in Jiangsu Province has been increasing, and the proportion of the distribution area of hotspot and sub-hotspot areas has risen from 19.6% to 45.8%, while the number of coldspot and sub-coldspot areas has decreased significantly. In terms of spatial distribution, the hotspot areas were concentrated in the southern part of Jiangsu Province in 2014, and by 2018, the rural tourism hotspot high-value agglomeration area began to move to the northwest, and in 2023, the hotspot areas further spread to the northeast. Overall, the distribution of hotspot zones has a tendency to shift from south to north, but the southern part of Suzhou is still an important inflow of rural tourism flows.

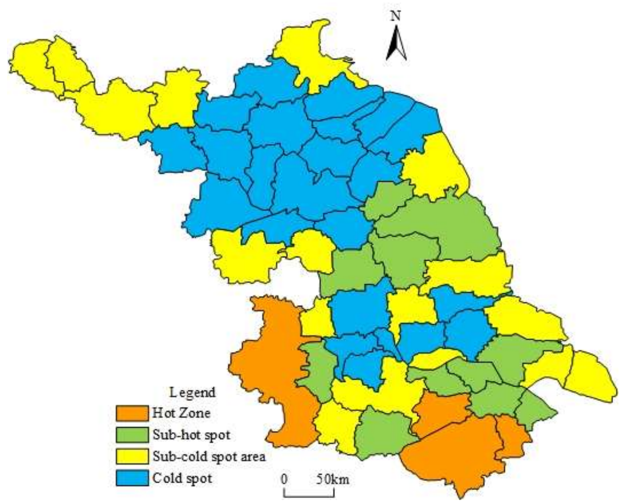
(2) Changes in the hotspot of counties

Yancheng City, Xinghua City and Dongtai City, Sheyang County have the greatest degree of change, jumping from sub-cold spot area and cold spot area to hot spot area and sub-hot spot area respectively.

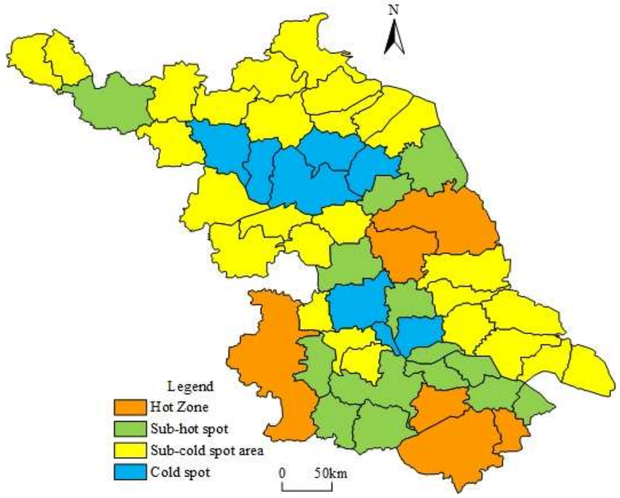
It can be found that counties and cities with a high level of economic and social development have a higher degree of rural tourism heat and a stronger diffusion effect on neighboring areas. Comparatively speaking, the weaker level of economic development in the northern part of Jiangsu Province, coupled with the constraints of ecological environment, climate and other natural factors, can't form a reciprocal position and advantage in the competition of rural tourism.




(a)2014




(b)2018




(c)2023

**Fig. 5.** The distribution map of 'hot spot' and 'cold spot' of rural tourism flow

# 4. Impact mechanism of rural tourism flows based on network data mining

According to the analysis of the spatio-temporal evolution characteristics of rural tourism flows, the exploration of the influencing factors of rural tourism flows is of great significance for promoting the development of rural tourism industry and enhancing the competitiveness and sustainable development of rural tourism destinations. In this chapter, we will continue to explore the influencing factors of rural tourism flows with Jiangsu Province of China as the study area.

## *4.1. Methodology for analyzing the mechanisms influencing the structure of tourism flows*

The QAP model is based on the social network analysis perspective, used to measure the existence of a certain spatial interaction between different "relationship" data, the model data are in the form of matrices.QAP regression technique is different from the general statistical methods, such as the common multiple regression analysis, in the face of the problem of multiple covariance, there are certain limitations. Unlike common statistical methods such as multiple regression, which have limitations when faced with the problem of multicollinearity and emphasize the need for variables to be independent of each other, the QAP method is mainly based on the "relationship" between the variables, and uses a matrix network as the explanatory and interpreted variables, which makes it more suitable for analyzing the influencing factors of spatial correlation. If there is no autocorrelation between the variables, the autocorrelation structure presents a monotonic function with coefficients Km, Kn, and the model data are generally presented as:

$$x_{ij} = \varepsilon_{ij} \tag{13}$$

$$y_{ij} = k_{yx} x_{ij} + \varepsilon_{ij} \tag{14}$$

$$k_{yx} = 0, y_{ij} = \varepsilon_{ij} \tag{15}$$

The data representation when the model produces autocorrelation is shown:

$$\varepsilon_{yij} = K_m v_{yi} + K_n v_{yi} + K_{ij} v_{yij} \tag{16}$$

$$\varepsilon_{xij} = K_m v_{xi} + K_n v_{xj} + K_{ij} v_{xij} \tag{17}$$

$$0 \le K_{ij} \le 1, K_m = 1 - K_{ij}, K_m = K_n \tag{18}$$

All other random variables independently obey the $N(0,1)$ distribution. While $K_m$ and $K_n$ represent the autocorrelation coefficients between the rows and columns of each matrix, respectively, $K_{ij}$ represents the coefficient of the random term in each statistical test. In the calculation of the complete QAP method, the above conditions are also required to be satisfied, and $K_n$ represents the overall autocorrelation coefficient. If $K_m$ is 0, then there is no autocorrelation among the matrices in the model, and if $K_m$ is 1, then there may be autocorrelation among the matrices in the model. The common explanatory variables set by comprehensive spatial correlation are the level of openness to the outside world, spatial geographic distance, urbanization level, industrial structure and so on. At the same time, considering the availability of data, focusing on the selection of regional geographic location, differences in the level of economic development, differences in population density, differences in the level of science and technology, differences in industrial structure and differences in the level of opening up to the outside world as the influencing factors of rural tourism in Jiangsu Province are included in the QAP regression model. Its basic QAP model is set as follows:

$$T = f(GD, IS, TL, PD, EC, OL) \tag{19}$$

$$Y_{\bar{y}} = \beta_0 + \beta_1 GD + \beta_2 IS + \beta_3 TL + \beta_4 PD + \beta_5 EC + \beta_6 OL + \mu \tag{20}$$

Among them, $T$ is the spatial correlation matrix of Jiangsu Province, which is used as an explanatory variable here, and $GD$ is the geographic proximity matrix, which is marked as “1” if the cities are geographically close to each other, and “0” if the cities are not adjacent to each other. $IS$ is the industrial structure difference matrix, the industrial structure is measured by the proportion of GDP accounted for by the tourism industry. $TL$ is the rural tourism expenditure difference matrix, tourism level is mainly measured by the proportion of regional “tourism expenditure financial expenditure”. $PD$ is the population density difference matrix, the population density is characterized by the region's “year-end population number administrative land area”. $EC$ is the difference matrix of economic development level, which is characterized by the per capita GDP of each region. $OL$ is a matrix of differences in the level of openness to the outside world, which is measured by the proportion of total imports and exports to regional GDP.

### *4.2. Analysis of factors influencing the structure of tourism flows*

4.2.1. Construction of the indicator system. The dependent variable is the size of tourists (the entry degree of tourism flow nodes). In order to analyze the influencing factors corresponding to different stages of tourism development, this study uses quantitative analysis to explore the factors that influence the structure of rural tourism flows. The independent variable is the factor used to explain the dependent variable, and in this study, climatic conditions, spatial proximity, tourism resource endowment, economic development level, hospitality service capacity, transportation accessibility, information exchange level, transportation accessibility, and support strength are selected as the independent variables to evaluate the impact of these factors on the change of tourists' scale, and the specific evaluation index system is shown in Table 3.

**Table 3.** Evaluation index system

| Index | Independent variable | Variable name |
|---|---|---|
| Climatic condition | Winter temperature | The average temperature in winter (°C) |
| | Summer temperature | The average temperature in summer (°C) |
| Geographic space | Space proximity | Is space adjacent |
| Tourism resource endowment | Tourist resources | 5A and 4A scenic spots weighted score |
| Economic development level | Industrial structure | The proportion of tertiary industry in GDP(%) |
| | Economic level | Gross regional product(billion yuan) |

| | | |
|---|---|---|
| Reception service capacity | Size of tourists | Tourists(ten thousand people) |
| | Travel income | Tourism revenue(billion yuan) |
| Communication level | Information communication | Network information search index |
| Traffic facilitation | Drive solo | The shortest time of the car(h) |
| | Railway travel | Shortest travel time by train(h) |
| Government support | Policy factor | Fiscal Expenditure / GDP(%) |

4.2.2. Analysis of the importance of influencing factors. According to the QAP regression analysis method, we regressed the relationship matrix of influencing factors related to rural tourism flow, and obtained the regression coefficients and test indexes of each variable matrix. Using the standardized coefficients for analysis, the QAP correlation coefficients between the variables in 2014, 2018, 2022 and 2023 were listed, as shown in Table 4. The regression analysis of rural tourism flow and the matrix of each influencing factor through the construction of the model shows that the direction and degree of the influence of different driving factors on the correlation network of rural tourism flow have changed differently over time, and have shown more obvious differentiation characteristics. The specific analysis is as follows:

1) Climate factors. The effects of climate factors are all positive, but most of them do not reach a significant level of more than 10%. Although the climate factor will have an impact on the structure of tourism flow to a certain extent, this effect is very limited. Specifically, the average winter temperature reaches a significant level only in 2018, which suggests that snow- and ice-themed tourism plays an important role in strengthening the spatial attractiveness of tourism flows. However, in the three years of 2014, 2022 and 2023, the average summer temperature does not pass the significance test of more than 10%, which obviously indicates that climatic factors cannot be the core factor influencing the structure of local tourism flows in Jiangsu Province.

2) Spatial proximity factor. The influence coefficient of geospatial proximity is significantly positive, indicating that geospatial proximity plays an important role in the influence of tourism flow structure. Inter-regional proximity promotes the formation of spatial association networks of tourism flows and generates spatial spillover effects in the spatial distribution of tourism flows. In terms of spatial spillover, this study notes that there is a stronger spatial spillover effect between neighboring prefecture-level cities, indicating that there is a more significant role of mutual promotion between tourism flows between neighboring regions. The reasons for this phenomenon may be related to factors such as transportation convenience, cultural inheritance and historical origin between neighboring regions.

3) Industrial structure. The influence coefficients of industrial structure experience changes from negative to positive, but none of the coefficients pass the 10% significance test, indicating that the influence of industrial structure on tourism flow is not obvious. It may be because the development of today's tourism is no longer limited only to the connection with a specific industry, but is more diversified and comprehensive. Comprehensive tourism is the current hotspot, and tourists are more inclined to diversified tourism experience, which leads to the relatively small influence of industrial structure on tourism flow.

4) Economic level. The coefficients of the influence of the economic level between regions are all significantly positive. The coefficients of the differences in regional GDP in 2022 are negative, while the coefficients of the other years are positive, but none of them pass the significance test. This indicates that the differences in regional economic development do not have a significant effect on the spatial correlation between regions.

5) Tourism Reception. The coefficients of the impact of tourism reception capacity show an overall positive trend, except for the year 2022. This means that the greater the difference in tourism numbers between regions, the more it promotes the spatial correlation of tourism flows, thus stabilizing the network of tourism flows between regions.

6) Tourism income. The impact coefficient of tourism income level shows a trend from negative to positive, and this change is proved to be significant after the significance test. The negative impact coefficient in the early stage indicates that the greater the difference in tourism income between regions, the weaker the connection between the regions in the tourism flow network and the smaller the strength of their association. And the positive impact coefficient in the later period indicates that the greater the difference in tourism income between regions, the stronger the link between regions in the tourism flow network and the stronger the strength of its association. It may be due to the fact that as the level of economic development increases, the tourism industry economy of each region has gradually developed, and the tourism flow network has more opportunities for cooperation and connection to form a closer network connection.

7) Information Exchange. The impact coefficients of information exchange are all significantly positive, and the coefficient values of the first three periods gradually increase over time, which to some extent indicates that with the development of network information technology, intelligent communication devices represented by mobile communication technology gradually play a role in the tourism links between the two regions.

8) Transportation conditions. The influence coefficients of transportation conditions are all significantly negative, indicating that the transportation conditions are negatively correlated with the intensity of tourism flow between the two places. This indicates that the distance decay law can well explain the travel pattern of tourists, i.e., there are more tourism links between regions that are close to each other. The coefficients for self-driving trips as well as railroad trips are significantly negative. This indicates that the more pronounced the spatio-temporal compression between the two regions, the stronger the tourism link between the two regions. It can be inferred from this result that self-driving and railroad have a significant role in promoting tourism.

9) Policy factors. The influence coefficient of the policy factor is positive in a few years. Specifically, the coefficient is positive in 2018, while it is negative in other years. The results indicate that the smaller the difference in the ratio of fiscal expenditure to GDP in 2022 and 2023, the weaker the tourism link between the two regions. Conversely, the greater the ratio of fiscal expenditure to GDP, then the stronger the tourism link between the two locations. It should be emphasized that this conclusion, although statistically supported, needs to be judged in context, and the impact brought about by the differentiation of policies in different regions on the behavior of tourists needs to be taken into account.

**Table 4.** QAP correlation coefficient

| - | 2014 | 2015 | 2022 | 2023 |
|---|---|---|---|---|
| Intercept | 0.000 | 0.000 | 0.000 | 0.000 |
| Winter temperature | 0.025 | 0.119*** | 0.04 | 0.034 |
| Summer temperature | -0.032 | 0.077 | 0.022 | 0.042 |
| Proximity space | 0.208*** | 0.228*** | 0.211*** | 0.258*** |
| Tourism resources | 0.182*** | 0.076** | 0.087** | 0.057* |
| Industrial structure | -0.085 | 0.043 | 0.032 | 0.038 |
| Economic level | 0.072* | -0.238*** | -0.155*** | -0.185*** |
| Tourism reception | 0.026 | 0.075 | -0.023 | 0.056 |
| Tourism income | -0.248*** | -0.096* | 0.111* | 0.104* |

| | | | | |
|---|---|---|---|---|
| Information exchange | 0.067** | 0.082** | 0.108*** | 0.062** |
| Self-driving travel | -0.005 | -0.077* | -0.123** | -0.138*** |
| Railway travel | -0.052 | -0.056* | -0.043*** | -0.026*** |
| Policy factors | -0.035 | 0.027 | -0.057** | -0.048* |

## 5. Conclusion

This paper takes rural tourism flow as the research object, selects Jiangsu Province of China as the research area, takes network data mining as the breakthrough, analyzes the spatio-temporal pattern of rural tourism flow and its evolution characteristics through hotspot measurement model, spatial analysis method, hotspot identification and other methods, and explores the influencing mechanism of the formation of rural tourism flow with the help of QAP model.

The analysis of the spatio-temporal evolution of rural tourism flows shows that the number of scenic spots in Jiangsu Province increases year by year, but the spatial difference is significant.The number of star-rated rural tourism scenic spots in Jiangsu Province grows from 19 in 2014 to 289 in 2023.The proportion of rural tourism scenic spots in the central and northern regions of Jiangsu grows to 47.26% in 2023, while the southern region of Jiangsu gathers the entire Jiangsu Province with a land area of 27.16% and 50.32% of rural tourism scenic spots. 50.32% of rural tourist attractions. This proves the positive correlation between the distribution of rural tourist attractions and the level of economic development. The center of gravity of the spatial distribution of rural tourist attractions generally moves to the north west direction, and the coverage of star-rated rural tourist attractions is expanding year by year. On the time evolution of rural tourism heat, the change trajectory of rural tourism heat in Jiangsu Province is basically consistent with the development stage of rural tourism, and the Gini coefficient of rural tourism heat increased from 0.51 in 2014 to 0.72 in 2018, and then fell back to 0.65 in 2023, presenting an "inverted U"-shaped evolutionary trajectory, which to a certain extent indicates that the rural tourism heat has a positive correlation with the economic development level. To a certain extent, it shows that the spatial difference in the degree of rural tourism hotness has shown a decreasing and weakening trend. The proportion of the distribution area of hot spots and sub-hot spots rises from 19.6% to 45.8%, and the number of cold spots and sub-cold spots shows a clear trend of reduction. The distribution of hot spots has a tendency to shift from the south to the north, but at present, it is still an important inflow of rural tourism flow in the south of Jiangsu Province.

Climatic conditions, spatial proximity, etc. are selected as independent variables, and tourist scale (inflow of tourism flow nodes) is used as the dependent variable to construct an indicator system to analyze the influence mechanism of rural tourism. The influence of climatic factors and inter-regional economic level are positive, and the influence coefficient of industrial structure experiences a change from negative to positive, but most of its coefficients do not pass the significance test. The influence coefficients of spatial proximity factor, tourism reception factor and information exchange factor are all significantly positive, which play an important role in the influence of tourism flow structure. The influence coefficient of tourism income factor shows a trend from negative to positive, and the change is significant. The influence coefficient of the transportation condition factor is significantly negative, which is negatively related to the intensity of tourism flow. The impact coefficient of the policy factor is positive only in a few years and is tested to be significant, indicating that the larger the ratio of fiscal expenditure to GDP, the stronger the tourism linkages between regions.

## References

[1] An, W., & Alarcón, S. (2020). How can rural tourism be sustainable? A systematic review. Sustainability,

12(18), 7758.

[2] Liu, Y. L., Chiang, J. T., & Ko, P. F. (2023). The benefits of tourism for rural community development. Humanities and Social Sciences Communications, 10(1), 1-12.

[3] Ćurčić, N., Mirković Svitlica, A., Brankov, J., Bjeljac, Ž., Pavlović, S., & Jandžiković, B. (2021). The role of rural tourism in strengthening the sustainability of rural areas: The case of Zlakusa village. Sustainability, 13(12), 6747.

[4] SUN, J. X., ZHANG, L. Y., & LUO, Y. L. (2023). Development of rural tourism resources in China under the goal of common prosperity: Current situation, problems and the development path. Journal of natural resources, 38(2), 318-334.

[5] Marzo-Navarro, M., Pedraja-Iglesias, M., & Vinzón, L. (2017). Key variables for developing integrated rural tourism. Tourism Geographies, 19(4), 575-594.

[6] Kataya, A. (2021). The impact of rural tourism on the development of regional communities. Journal of Eastern Europe Research in Business and Economics, 652463.

[7] DANILOSKA, N., & HADZI NAUMOVA-MIHAJLOVSKA, K. (2015). Rural tourism and sustainable rural development. Economic Development/Ekonomiski Razvoj, 17(3).

[8] Yang, J., Yang, R., Chen, M. H., Su, C. H. J., Zhi, Y., & Xi, J. (2021). Effects of rural revitalization on rural tourism. Journal of Hospitality and Tourism Management, 47, 35-45.

[9] Martínez, J. M. G., Martín, J. M. M., Fernández, J. A. S., & Mogorrón-Guerrero, H. (2019). An analysis of the stability of rural tourism as a desired condition for sustainable tourism. Journal of Business Research, 100, 165-174.

[10] Lv, L., Hu, J., Xu, X., & Tian, X. (2021). The evolution of rural tourism in wuhan: Complexity and adaptability. Sustainability, 13(24), 13534.

[11] Zhang, X. (2012). Research on the development strategies of rural tourism in Suzhou based on SWOT analysis. Energy Procedia, 16, 1295-1299.

[12] Germanovich, A. G., Vasilieva, O. N., Ordynskaya, M. E., Allanina, L. M., & Gorokhova, A. E. (2020). Impact of tourism on sustainable development of rural areas: International experience. Journal of Environmental Management and Tourism, 11(4), 965-972.

[13] Jiang, Q. (2021). Analysis of Rural Tourism Demand Characteristics and Experience Differences Based on Association Rule Mining. Wireless Communications and Mobile Computing, 2021(1), 8742950.

[14] Shen, S., Wang, H., Quan, Q., & Xu, J. (2019). Rurality and rural tourism development in China. Tourism Management Perspectives, 30, 98-106.

[15] Villanueva-Álvaro, J. J., Mondéjar-Jiménez, J., & Sáez-Martínez, F. J. (2017). Rural tourism: Development, management and sustainability in rural establishments. Sustainability, 9(5), 818.

[16] Qin, X., Li, X., Chen, W., Tan, H., Luo, L., & Xu, X. (2022). Tourists' digital footprint: the spatial patterns and development models of rural tourism flows network in Guilin, China. Asia Pacific Journal of Tourism Research, 27(12), 1336-1354.

[17] Li, Y., Gong, G., Zhang, F., Gao, L., Xiao, Y., Yang, X., & Yu, P. (2022). Network structure features and influencing factors of tourism flow in rural areas: Evidence from China. Sustainability, 14(15), 9623.

[18] Su, F., Duan, C., & Wang, R. (2020). Optimization model and algorithm design for rural leisure tourism passenger flow scheduling. Ieee Access, 8, 125295-125305.

[19] Li, T., Li, Q., & Liu, J. (2021). The spatial mobility of rural tourism workforce: A case study from the micro analytical perspective. Habitat International, 110, 102322.

[20] Chen, X., & Cong, D. (2022). Application of Improved Algorithm Based on Four-Dimensional ResNet in Rural Tourism Passenger Flow Prediction. Journal of Sensors, 2022(1), 9675647.

[21] Li, L., Tao, Z., Lu, L., Liu, H., Zhang, J., & Zhang, M. (2022). The impact of COVID-19 on the regional tourism flow network: An empirical study in Hubei Province. Current issues in Tourism, 25(2), 287-302.

[22] Xie, Y., Meng, X., Cenci, J., & Zhang, J. (2022). Spatial pattern and formation mechanism of rural tourism resources in China: Evidence from 1470 national leisure villages. ISPRS International Journal of Geo-Information, 11(8), 455.

[23] Wu, L., & Zhang, J. (2023). A study on the supply factors and mechanism of the inter-regional diffusion and transfer of inbound tourism flow. Current Issues in Tourism, 26(3), 392-406.

[24] Zhao Ziwei,Zhao Zuoquan & Zhang Pei. (2023). A new method for identifying industrial clustering using the standard deviational ellipse. Scientific Reports(1),578-578.

[25] Huang Peng,Miao Qiuhua,Sang Gao,Zhou Yuhang & Jia Minping. (2021). Research on quantitative method of particle segregation based on axial center nearest neighbor index. Minerals Engineering106716-.

[26] Zhigang Wang,Zhen Su,Ye Deng,Jürgen Kurths & Jun Wu. (2024). Spatial network disintegration based on kernel density estimation. Reliability Engineering and System Safety110005-.

[27] Laxmi Gupta& Jagabandhu Dixit. (2024). Spatial analysis of urban green space and its utilization rate for the flood-prone region Assam, India. Environment, Development and Sustainability(prepublish),1-31.